

(1)

الف) در شرایطی که تعداد صفحه بافر به اندازه کافی داشته باشیم از جذر مینیمم تعداد صفحات رابطه ها بیشتر باشد هر دو هزینه یکسانی دارند. هنگامی که داده ها به صورت uniform پخش نشده باشند استفاده از Hash Join مناسب نخواهد بود؛ همچنین هنگامی که نیاز داریم داده ها را به صورت مرتب داشته باشیم استفاده از Sort-Merge Join بهتر است.

ب) به اسلایدهای درس مراجعه شود

ج) ممکن است که هیچ شاخصی روی داده ها نداشته باشیم و مجبور به استفاده از full table scan باشیم حالت دیگر این است که شاخص unclustered روی داده ها داشته باشیم و اگر تعداد داده های پرس و جو زیاد باشد scan همه آنها به احتمال زیاد هزینه کمتری دارد. حالت دیگری نیز وجود دارد و این است که شاخص داشته باشیم ولی شرط دیگری وجود داشته باشد که برای آن نیاز به اسکن کل داده ها داشته باشیم مثلاً این شرط به صورت or با attribute ای باشد که روی آن شاخص داریم

(2)

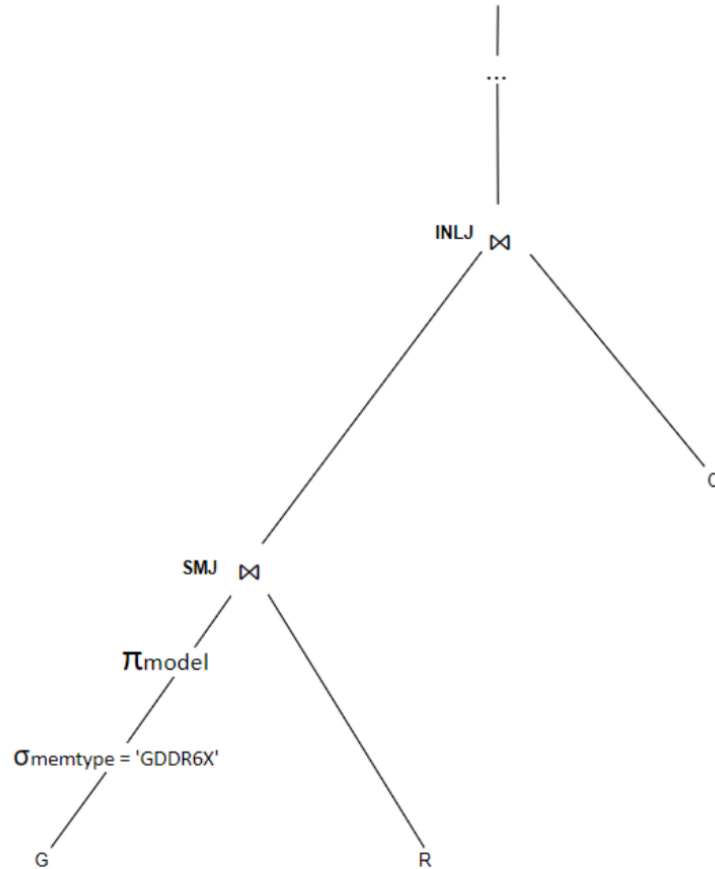
الف) هر صفحه میتواند $40 = 4096/100$ تاپل را نگه دارد. بنابراین تعداد صفحات مورد نیاز برای 100000 تاپل $2500 = 100000/40$ صفحه است.

ب) مطابق بخش الف میتوان بدست آورد که S شامل 981 صفحه است. از آنجایی که با توجه به تعداد بافر موجود، مرتب کردن هر کدام از رابطه ها نیاز به 3 pass دارد. بنابراین در کل هزینه $2 * 3 * 2500$ برای مرتب سازی R و $2 * 3 * 981$ برای مرتب سازی S می دهیم و با $981 + 2500$ نیز جویین را انجام می دهیم.
هزینه کل:

$$24367 = 3481 + 5886 + 15000$$

ج) هزینه Hash Join برابر $3 * (N + M)$ می باشد که در این صورت هزینه برابر $3 * (981 + 2500) = 10443$

(3) پلن نهایی به شکل زیر می‌شود



محاسبه هزینه:

با استفاده از داده های مسئله میدانیم در C و R و G به ترتیب ۸۰ و ۴۵۰ و ۴۰ پیج داریم. ابتدا هزینه SMJ را محاسبه میکنیم
ابتدا G را میخوانیم و سلکشن و پروجکشن را اجرا می‌کنیم تعداد پیچها ۴۰ تاست که بعد از سلکشن میشود ۴ پیج که بعد از پروجکشن میشود ۱ پیج بدون هیچ I/O ای مرتب میشود
R هم ۴۵۰ پیج است که در ۳ pass مرتب می‌شود همچنین لازم است که نتیجه را با هم merge کنیم

$$\text{Cost} = 40 + 1 + 2 * 3 * 450 + 450 + 1 = 3,192$$

حال باید تعداد تاپل های حاصل SMJ را محاسبه کنیم. چون توزیع uniform است هر G با ۱۵ تاپل از R جوین میشود. بعد از سلکشن تعداد تاپل های G برابر ۱۲۰۰ میشود پس خروجی این جوین ۱۸ هزار تاپل خواهد بود.

چون رابطه داخلی C است و جویین روی ssh که Primary Key این رابطه است انجام میشود به ازای هر کدام از این ۱۸ هزار تاپل یک تاپل از C جویین میشود

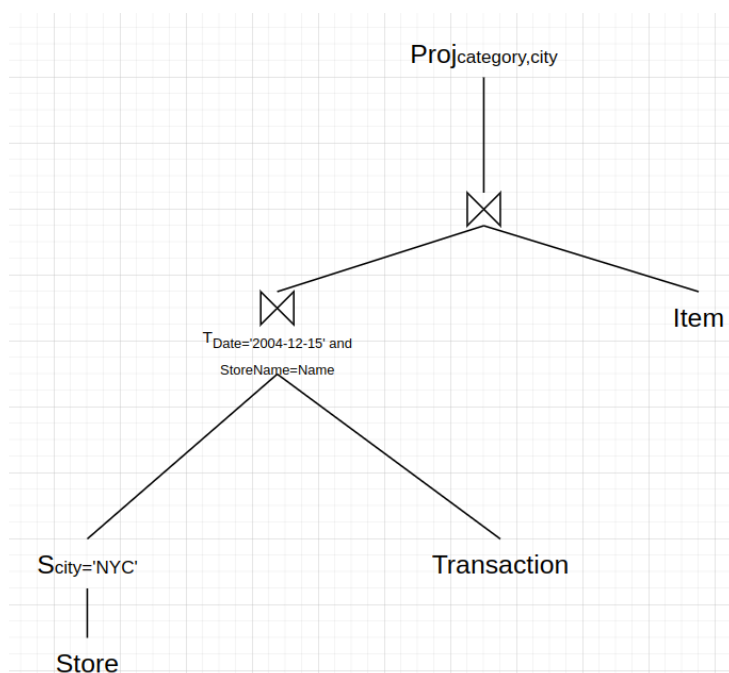
$$\text{Cost} = 18000 * (3 + 1) = 72,000$$

سپس باید بخاطر دستور Group by رکوردها را مرتب کنیم که رکوردها از پیش به علت SMJ مرتب شده هستند.

$$\text{Total Cost} = 3,192 + 72,000 = 75,192$$

همچنین دقت شود که از REFINED SMJ در این راه حل استفاده نشد.

(4) پلن نهایی به شکل زیر است.



$S_{city='NYC'}$: use B+ Tree index -> select 10 tuples -> 2 pages -> cost = 2 + 2 = 4

Store Join Transaction: use INLJ -> cost = 10 * 3 = 30 (after selection 100 tuple remains)

Join with Item: use unclustered hash index on name -> cost = 100 * (1.2 + 1) = 220

Total cost = 4 + 30 + 220 = 254 I/O